Exploratory Research Proposal (FY 1999)

The Role of RNA Editing Processes: A Simulation Approach

Technical Category: Biosciences

Principal Investigator: George Papcun, CIC-3, MS P990, gjp@lanl.gov, tel: 7-9800

Senior Investigator: Luis Rocha, CIC-3, MS P990, rocha@lanl.gov, tel: 5-5328

Funding Requested: $200,000/year for a period of two years

Abstract

The discovery of messenger RNA (mRNA) molecules containing information not coded in DNA, first persuaded researchers in molecular biology that some mechanism in the cell might be responsible for post-transcriptional alteration of genetic information; this mechanism was called 'RNA Editing' [Benne at al, 1986]. “It was coined to illustrate that the alterations of the RNA sequence (i) occur in the protein-coding region and (ii) are most likely the result of a post-transcriptional event” [Benne, 1993, page 16]. The term is used to identify any mechanism which will produce mRNA molecules with information not specifically encoded in DNA. Initially, the term referred to the insertion or deletion of particular bases (e.g. uridine), or some sort of base conversion (e.g. adenosine → guanisine). Today, more RNA editing mechanisms, have been observed [for a good review please refer to [Arts and Benne, 1996] .

“In spite of the diversity and (presumed) differences in mechanism, research on different forms of RNA editing addresses the same questions: (1) what are the *cis*-acting RNA elements that designate a certain site for sequence alteration, (2) what are the *trans*-acting factors that operate in editing reactions and what is their mechanism of action, and (3) why do RNA editing processes exist?” [Ibid, pp. 39-40]. This LDRD exploratory research (ER) project aims at investigating the third question from a complex, evolutionary, systems perspective. In particular we aim at:

1. Understanding the ***theoretical nature of*** genetic systems with RNA Editing, drawing from concepts developed in the fields of complex systems, theoretical biology, and bio-semiotics.
2. Developing ***computer simulations*** of populations of artificial organisms with different RNA editing capabilities, in order to investigate the hypotheses that RNA editing provides an evolutionary advantage to some organisms or that it is a relic of an ancient RNA world.

2. Background and Scientific Impact

The most famous RNA editing system is that of the African Trypanosomes [Benne, 1993; Stuart, 1993]. Its genetic material was found to possess strange sequence features such as genes without translational initiation and termination codons, frame shifted genes, etc. Furthermore, observation of mRNA's showed that many of them were significantly different than the genetic material from which they had been transcribed. These facts suggested that mRNA's were edited post-transcriptionally. It was later recognized that this editing was performed by *guide* RNA's (gRNA's) coded mostly by what was previously thought of as non-functional genetic material [Sturn and Simpson, 1990; Blum, Bakalara, and Simpson, 1990]. In this particular genetic system, gRNA's operate by inserting, and sometimes deleting, uridines. To appreciate the effect of this edition consider figure 1. The first example [Benne, 1993, p. 14] shows a massive uridine insertion

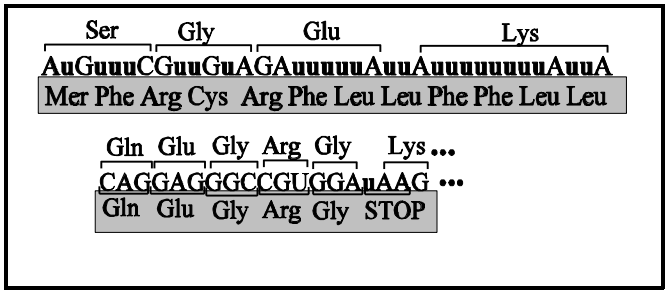

Figure 1: U-insertion in Trypanosomes' RNA (from [Rocha, 1995])

(lowercase u's); the aminoacid sequence that would be obtained prior to any edition is shown on top of the base sequence, and the aminoacid sequence obtained after edition is shown in the gray box. The second example shows how potentially the insertion of a single uridine can change dramatically the aminoacid sequence obtained; in this case, a termination codon is introduced. It is important to retain that a mRNA molecule can be edited in different degrees precisely according to the concentrations of editing operators it encounters. Thus, at the same time, several different proteins coded by the same gene may coexist, if all (or some) of the mRNA's obtained from the same gene, but edited differently, are meaningful to the translation mechanism.

The role of RNA editing in the development of more complex organisms has also been shown to be important, Lomeli at al [1994] have discovered that the extent of RNA editing affecting a type of receptor channels responsible for the mediation of excitatory post-synaptic currents in the central nervous system, increases in rat brain development. As a consequence, the kinetic aspects of these channels will differ according to the time of their creation in the brain's developmental process. Similar RNA editing processes have been identified in mammalian brains [Simpson and Emerson, 1996], including human brains [Mittaz et al, 1997]. Such editing may be involved in neurological diseases such as epilepsy [Ibid]. The hepatitis B virus is also responsible for insertional editing of human genes presumed to play a role in hepatocarcinogenesis [Graef et al, 1994]. RNA editing has also been encountered in the aberrant edition of encoded liver proteins, which could contribute to potent liver oncogenesis in rats [Yamanaka, 1997]. Conversely, RNA editing may offer a fruitful avenue for gene therapy, as it could be used to destroy specific, unwanted, mRNA's [Kozarsky and Couture, 1997].

Clearly, then, RNA editing is proving to be a key factor in some genetic systems, and though more and more knowledge is accumulated to answer Arts and Benne's [1996] first two questions quoted in the abstract, not much has been advanced to respond to the third question: "Why do RNA editing processes exist?". Two main hypothesis exist: the *selective* and the *neutralist* hypothesis [Ibid]. The first posits that RNA editing offers an evolutionary advantage in the regulation of gene expression and the second that it is might be a relic of an ancient RNA world. We believe we can advance some answers to this question by framing the problem in a more theoretical framework, and developing computational simulations to test aspects of these hypothesis.

2. The Semiotics of RNA Editing: A Theoretical Model

Semiotics concerns the study of signs/symbols in three basic dimensions: syntactics (rule-based operations between signs within the sign system), semantics (relationship between signs and the world external to the sign system), and pragmatics (evaluation of the sign system regarding the goals of their users) [Morris, 1946]. We can understand the semiotics of the genetic system if we consider all processes taking place before translation as the set of syntactic operations; the relation between mRNA (signifier) and folded amino acid chains (signified), through the genetic code, as the implementation of a semantic relation; and finally, the selective pressures on the developed phenotypes as the pragmatic evaluation of the genetic sign system. Figure 2 depicts these relationships.

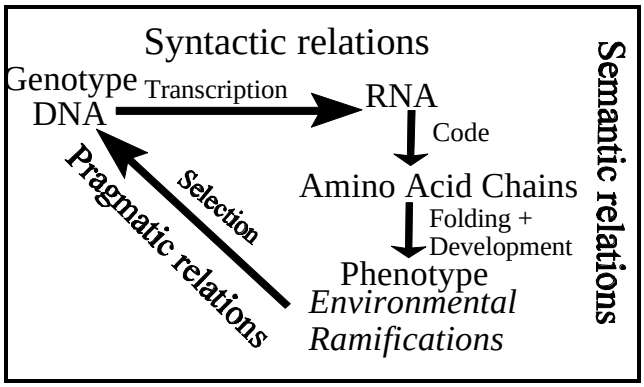

Figure 2: Semiotics of the Genetic System (adapted from Cariani [1998])

Until now, the semiotics of DNA has been considered to be strictly unidirectional: DNA stands for proteins to be constructed. In other

words, the symbolic DNA encodes (through the genetic code) phenotypes with repercussions in some environment. If in addition to symbols standing for actions to be performed, the genetic sign system is also allowed a second type of symbols standing for environmental, contextual, measurements, then a richer semiotics can be created which may have selective advantage in rapidly changing environments, or in context dependent, developmental processes. Figure 3 depicts such a sign system. The top plane contains two different types of symbols which are combined in different ways (symbolic operations). *Type 1* symbols stand for actions through a *code* ϕ (e.g. the genetic code) and *type 2* symbols stand for measurements through a different *code* γ which is being hypothesized here[1].

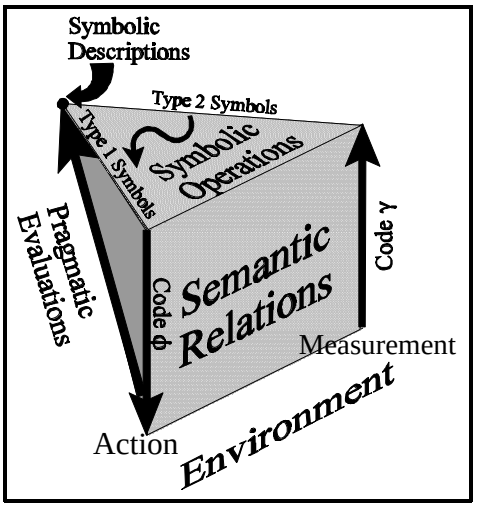

Figure 3: Genetic semiotics with 2 type symbol system

If a second type of symbols exists, which operate with genetic messages and in so doing change the latter's encoded meaning, their access to environmental information can provide the genetic system real-time control of genetic expression according to context. This ability would certainly be useful for phenotypical development in changing environments. Some evidence has been presented [Benne, 1993; Stuart, 1993; Simpson and Maslov, 1994; Lomeli et al, 1994] that RNA Editing is used in some genetic systems in different amounts according to different contexts (namely, different stages of a developmental process).

Indeed, if the concentrations of editing operators in a genetic system with RNA editing can be linked to environmental information, the concentrations of different proteins obtained may be selected accordingly, and thus evolve a system which is able to respond to environmental changes without changes in the major part of its genetic information (genome size optimization). One gene, different contexts, different proteins. This may be precisely what the Trypanosome parasites have achieved: control over gene expression in different parts of their life cycles.

More interestingly, RNA editing may be more than just a system responsible for the introduction of uncertainty (one-to-many relations), but also, and paradoxically, a system that may allow the evolution of different proteins constrained by the same genetic string. In other words, even though one gene may produce different mRNA's (and thus proteins), the latter are not allowed heritable variation since they are always constrained by the gene from which they are edited, and which is ultimately selected and transmitted to the offspring of the organism. We can see RNA Editing, especially in the case of gRNA's, as a case of co-adaption of two distinct systems: the stored genetic information and the contextual editors, also stored in DNA, but independent and meaningless to the larger semantic loop of the genetic code.

Benne [1993, p. 22] has shown that the dependent evolution of one gene and several contexts may allow the introduction of highly specific, local (contextual) changes, more effectively than the independent evolution of several genes. If all of the different expressions were allowed different genes, they would evolve separately not only increasing the size of the genome, but also, possibly, making it harder to maintain coherent, multicellular, phenotypes as well as coherent developmental processes. For instance, the editing of several genes of the Trypanosoma Brucei is developmentally regulated [Stuart, 1993] which may be of evolutionary advantage for these parasites [Simpson and Maslov, 1994]. Though in the course of evolution editing was partially or completely eliminated in many lineages of eukaryotic organisms containing

[1] For more details about the semiotics of genetic 2 type symbol systems, please refer to [Rocha, 1995, 1997, 1998]. Notice that code γ is proposed here as an abstraction referring to the set of mechanisms which will link environmental measurements (context) to *type 2* symbols. It is not expected to function as a proper genetic code with clear cut symbols (nucleotide codons standing for aminoacid chains). In simple terms, what I refer to as a code here is any mechanism able to relate "inert" material structures (signifiers) to other material structures with some functional dynamics (signifieds) in some organism/environment coupling.

mithocondria, by reverse transcription of partially edited mRNA's, it may be useful for the development of parasitic adaptations as is the case of the developmental regulation of editing in T. Brucei, because parasites need to survive in several completely different environments which require very different responses from them[Ibid]. The African Trypanosomes for instance, use the famous Tsé Tsé flies as carriers before infecting mammals; both present the parasite with completely different environments that trigger in it very different stages of development, at least in great part through the workings of the RNA editing system.

We can thus think of DNA-based genetic information as a set of symbolic descriptions based on two types of symbols: *type 1 symbols* expressed in mRNA molecules and standing for actions to be performed; *type 2 symbols* expressed in gRNA molecules (or other editing mechanisms) and standing for contextual observables. RNA editing can be seen as a set of symbolic operations performed with symbols of both types, resulting in symbols of type 1 to be translated into actions by the genetic code. This implements the two type symbol semiotics system described above.

3. *Contextual Genetic Algorithms* and the Evolutionary Hypothesis

In Rocha[1995, 1997] a formal description of a computational genetic algorithm [Holland, 1975] was presented. The new algorithm, called *Contextual Genetic Algorithm*, provides a computational means to simulate a genetic system with editing (insertion/deletion) characteristics based on populations of editing elements with varying concentrations. This algorithm, which spawns from the semiotic framing of the RNA editing system as described in 2, can be used to test a multitude of evolutionary scenarios. We propose the simulation of several populations of artificial organisms, some endowed with RNA editing abilities, and some without. The artificial organisms are to be deployed in artificial environments with varying survival demands, or fitness landscapes. Assuming, in some cases, that the changing environment will trigger different concentrations of editing agents, we wish to investigate the evolutionary conditions under which the existence of an RNA editing system (or type 2 semiosis) may provide an evolutionary advantage in changing environments. We intend to set up experiments to test, for instance, different timings of the alternation of fitness landscapes, in a way using Levins' [1968, chapter 2] strategies of adaptation, here employed in the co-adaptation of genetic information with the editing system to a changing environment.

In biological genetic systems RNA editing regulates gene expression. Somehow, perhaps organisms have used the edition of mRNA molecules to their advantage by linking it to environmental context. If a particular external event has the effect of changing the concentrations of editing agents in some genetic system, then those genes which are able to produce fit phenotypes in the different contexts will be selected. Notice that changing environmental context will not merely affect the concentration of editing agents, but also, potentially, the fitness landscape of the genetic system. Thus, the ability to link changes in the environment with internal parameters such as concentrations of editing agents, can potentially give organisms an adaptive advantage as gene expression can become contextually regulated. The idea is the introduction of the second kind of semantic relation leading to a second type of symbol described in section 2. The computer simulations we propose, can experiment with such a hypothesis and thus shed some light on function of the RNA editing system.

4. *RNA Toy Worlds* and the Neutralist Hypothesis

The idea that life may have originated from pure RNA world has been around for a while [Eigen, 1992; Schuster, 1995]. In this scenario the first life forms relied on RNA molecules as both symbolic carriers of genetic information, and functional, catalytic molecules. The neutralist hypothesis for the function of RNA editing assumes such a RNA world origin of life. It posits that RNA editing could offer a process by which the dual role of RNA molecules as information carriers and catalysts could be more easily co-exist. The key problem for the RNA world origin of life hypothesis is precisely the separation between these two functions of RNA. On the one hand RNA molecules should be stable (non-reactive) to carry information, and on the other hand they should be reactive to perform their catalytic function. RNA editing, could be seen as means to fragment genetic information into several non-reactive molecules, that are later, through RNA editing processes, integrated into reactive molecules [Arts and Benne, 1996]. In this view, RNA editing now exists as a relic from a RNA world origin of life.

Schuster [1995, 1997] has developed several simulations of a RNA toy world to prove the principle of *shape space covering*: he has proved that only a small fraction of sequence space has to be searched in order to find a sequence that folds into a predefined structure. In the simulations we propose, we assume a set of predefined structures to be reactive RNA sequences. This set is a subset of the space of folded RNA sequences. Conversely, we also assume a set of non-reactive RNA sequences. Given many random distributions of the reactivity of a RNA sequence space, we wish to study how easily can reactive sequences be constructed from RNA edition of non-reactive molecules. We intend to use Reidys [1996] *random graphs* to perform these simulations. We thus wish to investigate the necessary conditions (distribution of reactivity) in toy RNA sequence spaces that would make the neutralist hypothesis for RNA editing feasible.

5. Expected Results

We expect to determine the conditions under which the hypothesis of sections 3 and 4 could be validated. We anticipate that both hypothesis, under a certain set of conditions, might be true. Indeed, as Gould [1995] likes to stress, evolution is much more opportunistic than we tend to think, and it is prone to accumulate historical contingencies. RNA editing could easily be an ancient historical contingency, that nonetheless may be still be used advantageously in present day DNA-Protein biology. Therefore, the two hypotheses are not incompatible. RNA editing may be both a relic of an RNA world, but given that it does exist in today's DNA-Protein life forms, it could very well be used advantageously in some set of circumstances. We expect our research to shed some light into the nature of those circumstances. The results of our simulations to determine the role of RNA editing in biology, will have implications for current research on such topics as gene therapy and RNA editing caused oncogenesis and could potentially open fruitful research avenues in these areas.

References

Arts, G.J. and R. Benne [1996]."Mechanism and evolution of RNA editing in kinetoplastida." *Biochimica et Biophysica Acta.* 1307, 39-54.

Benne, R. (Ed.) [1993]. *RNA Editing: The Alteration of Protein Coding Sequences of RNA.* Ellis Horwood Publishers.

Benne, R., J. van den Burg, J.P.J. Brakenhoff, J.H. van Boom, and M.C. Tromp [1986]."Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA." *Cell.* Vol. 46, pp. 819-826.

Blum, B., N. Bakalara, and L. Simpson [1990]."A model of RNA editing in kinetoplast mitochondria: "guide" RNA molecules transcribed from minicircle DNA provide the edited information." *Cell.* Vol. 60, pp.189-198.

Cariani, Peter [1995]."Towards an evolutionary semiotics: the role of symbols in organisms and adaptive devices." In: *Evolutionary Systems: Biological and Epistemological Perspectives on Selection and Self-Organization*. S. Salthe, G. Van de Vijver, and M. Delpos (eds.). Kluwer Academic Publishers.

Eigen, M. [1992]. *Steps Towards Life: A Perspective on Evolution.* Oxford University Press.

Gould, S.J. [1995]."The pattern of life's history." In: *The Third Culture: Beyond the Industrial Revolution*. J. Brockman (ed.). Simon and Schuster, pp. 51-73.

Graef, E., W.H. Caselmann, J. Wells, and R. Koshy [1994]."Insertional activation of mevalonate kinase by hepatitis B virus DNA in a human hepatoma cell line." *Oncogene.* Vol. 9, pp. 81-87.

Kozarsky, K.F. and L.A. Couture [1997]."Message therapy: gene therapy that tarhets mRNA sequence and stability." *American Journal of Human Genetics.* Vol. 61, pp. 790-794.

Levins, R. [1968]. *Evolution in Changing Environments: Some Theoretical Explorations.* Princeton University Press.

Lomeli, H. et al [1994]."Control of kinetic properties of AMPA receptor channels by RNA editing." *Science.* Vol. 266, pp 1709-1713.

Mittaz, L., S.E. Antonarakis, M. Higuichi, and H.S. Scott [1997]."Localization of a novel human RNA-editing deaminase (hRED2 or ADARB2) to chromosome 10p15." *Human Genetics.* V. 100, pp. 398-400.

Morris, C.W. [1946]. *Signs, Language, and Behavior.* G. Braziller Publishers..

Pollack, R. [1994]. *Signs of Life: The Language and Meanings of DNA.* Houghton Mifflin.

Reidys, C. [1997]. "Random graphs and sequence structure maps". *Combinatorics, Probability and Computing*. (In Press)

Rocha, Luis M. [1995]."Contextual genetic algorithms: evolving developmental rules." In: *Advances in Artificial Life*. J. Moran, A. Moreno, J.J. Merelo, and P. Chacon (Eds.). Springer -Verlag, pp. 368-382.

Rocha, Luis M. [1997]. *Evidence Sets and Contextual Genetic Algorithms: Exploring Uncertainty, Context and Embodiment in Cognitive and biological Systems.* PhD. Dissertation. State University of New York at Binghamton.

Rocha, Luis M. [1998]."Selected self-organization and the Semiotics of Evolutionary Systems." In: *Evolutionary Systems: Biological and Epistemological Perspectives on Selection and Self-Organization*. S. Salthe, G. Van de Vijver, and M. Delpos (eds.). Kluwer Academic Publishers, pp. 341-358.

Schuster, P. [Genotypes and phenotypes; adventures in an RNA toy world]."1997." *Biophysical Chemistry.* In Press. Available as a pre-print of the Santa Fe Institute series (97-04-036).

Schuster, P. [1995]."Artificial life and molecular evolutionary biology." In: *Advances in Artificial Life*. F. Moran, A. Moreno, J.J. Merelo, P. Chacon. Springer, pp. 3-19.

Seiwart, S.D. and K. Stuart [1994]."RNA Editing: transfer of genetic information from gRNA in Vitro." *Science.* Vol. 226, pp. 114-116.

Simpson, L. and R.B. Emerson [1996]."RNA editing." *Annual Review of Neuroscience.* V. 19, pp. 27-52.

Simpson, L. and D. Maslov [1994]."RNA editing and the evolution of parasites." *Science.* Vol. 264, pp. 1870-1871.

Stuart, K. [1993]."RNA editing in mitochondria of african trypanosomes." In: *RNA Editing : The Alteration of Protein Coding Sequences of RNA*. R. Benne (Ed.). Ellis Horwood Publishers. pp. 26-52..

Sturn, N.R. and L. Simpson [1990]."Kinetoplast DNA minicircles encode guide RNA's for editing of cytochrome oxidase subunit III mRNA." *Cell.* Vol. 61, pp.879-884.

Yamanaka, S.et al [1997]."A novel translational repressor mRNA is edited extensively in livers containing tumors caused by the transgene expression of the apoB mRNA-editing enzyme." *Genes and Development.* Vol. 11, pp. 3321-333.